利用 AI 深度伪造技术的涉网犯罪防治研究
——以“AI 去衣”技术为例

■ 黄炯量

摘　要　随着人工智能技术的快速发展，AI 深度伪造技术逐渐浮出水面，并逐渐被犯罪分子用于非法途径，对社会风气和道德观念都造成了极大的冲击，目前与 AI 深度伪造技术有关的涉网犯罪还在持续上升，亟需进行有效且系统的整治。本文以“AI 去衣”此种 AI 深度伪造技术为切入口，通过研究此类 AI 伪造技术的发展现状、犯罪特点、社会危害以及应对策略，以深入研究 AI 深度伪造技术的涉网犯罪的综合防治。旨在为相关部门提供有益的参考和借鉴，以助力构建一个安全、健康、有序的网络环境。

关键词　“AI 去衣”　伪造技术　涉网犯罪　防治对策

随着 AI 技术的快速发展，其在各个领域的应用也日益广泛。然而，技术的双刃剑效应也日益凸显，如 AI 深度伪造技术逐渐被非法使用，特别是在图像处理领域，“AI 去衣”伪造技术的出现给社会带来了极大的困扰和危害。这种技术通过算法对图像进行处理，在图片的基础上，通过 AI 自行运算使用相关的淫秽图像素材对图片进行重绘，从而生成裸体图像，被不法分子用于制作和传播淫秽物品，严重侵犯了公民的隐私权和人格尊严，对社会风气和道德观念造成了极大的冲击。

一、“AI 去衣”伪造技术的犯罪发展现状

数据显示，2023 年基于 AI 的深度伪造欺诈暴增了 3000%，全球 Deepfake 攻防挑战赛出题人、ZOLOZ 技术总监姚伟斌在接受媒体采访时表示，AIGC 的生成伪造，今年的攻击量比去年大概增加了 10 倍，整个风险其实是呈指数级上升的情况，不仅仅是在中国，在韩国、菲律宾、印尼等国家都有类似的情况发生。据网络安全公司 Home

作者：浙江警察学院侦查系

Security Heroes2023 年发布的数据，2023 年网络上可监测到的 AI 去衣图片就不低于 10 万张。在 2023 年 1 月，荷兰 DeepTrace 实验室发布了一份 Deep-Fake 发展报告，数据显示 2023 年“deepfake”关键词的谷歌搜索次数比 2022 年增长了近 100 倍，可见“AI 去衣”技术正在被越来越多人使用。

随着“AI 去衣”技术的发展，越来越多的人未经他人明确同意而使用他人照片，非法生成裸体图像，用于盈利等非法用途。以我国公开的一起“AI 去衣”的犯罪事件为例：在 2023 年 6 月至 8 月间，白某某在北京市海淀区等地，通过互联网发布“AI 一键去衣”广告，将他人提供的承载人脸信息的不特定多数女性图片，通过 AI 软件制作成裸体图片贩卖，同时出售“AI 一键去衣”软件及使用教程牟利。经查，被告人白某某通过社交软件将他人提供的女性图片（包括女明星、同事和同学的照片）制作成裸体图片并进行贩卖。总共贩卖“去衣”图片 6000 余张，其中 1500 余张图片被认定为淫秽物品。最终被告人白某某以牟利为目的，制作、贩卖淫秽物品，被以制作、贩卖淫秽物品牟利罪追究刑事责任。另外，白某某还因违反国家规定处理个人信息的行为，侵害了不特定多数公民的个人信息安全，损害了社会公共利益，承担了相应的民事侵权责任。

但由于“AI 去衣”技术发展的迅速性和新颖性，立法和侦查仍处在滞后阶段，相关法律和防治策略没有被系统的确立，有相当一部分的违法犯罪者并未得到应有的法律惩罚。目前我国可公开引用的判例也仅有白某某案这一例，且仅能根据刑法中的“传播和贩卖淫秽物品”等传统罪名进行定罪规制，并没有针对 AI 伪造的特定法律。近年来，在暗网、深网上此类犯罪更是越发猖獗，据数据统计，有接近七成的好莱坞女明星的“AI 去衣”图被制作发布在暗网上，且部分在暗网的非法图片已在社交媒体上大幅扩散，造成极其恶劣的影响。例如在 2024 年 1 月，由“AI 去衣”技术生成的明星泰勒 · 斯威夫特“不雅照”在 X（原推特）、脸书等社交媒体上流传，浏览量已过千万等。综上可见，此类涉网犯罪当前正处在高发态势，亟需相关部门进行有效的规制。

二、“AI 去衣”伪造技术的犯罪特点与社会危害

“AI 去衣”伪造技术是基于深度学习算法和图像处理技术的一种深伪（deepfake）应用，其核心是使用生成对抗网络（GANs），即通过学习大量真实人体图片数据，精准重建目标图像的纹理和细节。

（一）犯罪特点

1．犯罪技术门槛低

近年来，随着开源深度学习框架（如 TensorFlow 和 PyTorch）的普及，以及预训练模型和教程的广泛传播，犯罪分子可轻松利用现成工具实施“AI 去衣”犯罪，无需深厚的技术背景。例如，某些论坛和暗网社区甚至提供“一键生成”工具，普通用户只需上传目标图片，便可获得伪造图像。

2．犯罪成本低廉

由于“AI 去衣”伪造技术的自动化程度高和可定制性强，不法分子可以利用公开的软件或代码直接在线上进行图像伪造，无需借助复杂的工具和人力资源，使得该伪造过程几乎没有犯罪成本。

3．犯罪隐蔽性强

生成的“AI 去衣”图像往往通过社交

媒体、论坛或暗网传播。这种传播模式不仅速度快，而且高度匿名，给执法部门的监控带来巨大挑战。且有相当一部分的犯罪分子利用虚拟专用网络（VPN）、加密通信工具以及海外服务器发布内容和伪造图像，进一步增加了执法部门取证和追踪的难度。

4．犯罪后果易于传播

由于“AI 去衣”的技术已被应用于众多软件，形成系统化、自动化的制作流程。犯罪分子可以直接在相关软件中使用“一键生成”按钮，进行大规模的制作非法图片，制作过程快速且方便。且通过社交媒体和网络平台，伪造的淫秽图片能够迅速传播，造成广泛影响。

5．犯罪可定制性强

由于“AI 去衣”伪造图片可对受害人造成恶劣的社会影响，且会侵犯到其的人格尊严权。因此犯罪分子常针对目标人物（如公众人物、女性或青少年以及受人委托的指定人员）进行定向伪造，用以敲诈勒索或报复性传播，对特定人群造成不可挽回的伤害。

（二）社会危害

1．侵犯个人隐私

“AI 去衣”技术的滥用会直接威胁到受害者的个人隐私，伪造的裸露图像往往被用作网络性骚扰或敲诈工具，使受害者遭受心理创伤甚至社会群体的谴责和孤立，给其的个人生活和工作带来极大的困扰和不便。

2．社会信任危机

随着深伪技术的泛滥，公众会对图像真实性的信任逐步下降，从而趋向相信“任何图像”都可能会是被伪造的。这不仅威胁到个人和机构的声誉，也使媒体报道和司法证据的可靠性遭受质疑。

3．助长网络暴力

伪造内容一旦被不加辨别地传播，可能引发大规模网络暴力，对受害者及其家庭造成毁灭性影响。例如，2019 年曝光的“DeepNude”软件使数万女性的“AI 去衣”伪造图被大量泄露传播，使得许多受害者都暴露在公众的网络暴力之下，一时间严重破坏了社会风气。

4．法律和伦理困境

当前法律对深伪技术的规制普遍滞后，特别是“AI 去衣”技术所涉及的道德和法律问题复杂，司法部门难以找到具体法条作为定罪的根据，使得“AI 去衣”技术导致的伦理道德问题也难以从法律层面得到全面有效的解决。

三、AI 深度伪造涉网犯罪的防治策略与措施

（一）完善 AI 领域的法律法规，加强司法解释指导工作

就目前而言，在 AI 领域的立法速度已赶不上 AI 技术发展的速度，很多领域的 AI 技术都亟需特殊法的规制，故推动法律的立改废释纂，以良法促善治，提高依法治理水平应是首要任务。这需要针对“AI 去衣”伪造技术的犯罪特点和社会危害，加快完善相关法律法规的建设，明确界定该技术的犯罪行为、法律责任界定以及处罚标准。

在国内方面，我国最高人民法院在 2022 年 12 月 9 日发布了《关于规范和加强人工智能司法应用的意见》以落实《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》和《新一代人工智能发展规划》的具体举措，其中主要包括人工智能司法应用需要遵循的基本原则、人工智能司法应用的主要应用范围、

人工智能司法应用系统建设要求以及如何建设较为完备的司法人工智能技术应用和理论体系等意见，可见当前我国立法机关已经在有意强化在 AI 领域的法律规范，特别是在构建 AI 应用系统建设方面，明确了 AI 司法应用按照人民法院信息化建设发展的规划部署，并要求未来 AI 使用领域设计更为完善的信息系统架构和技术标准体系，体现出最高院对 AI 犯罪防治的立法侧重点及趋势，并侧面展现了我国未来 AI 规范治理步调。对比国际方面，欧盟于 2024 年正式通过了全球首部《人工智能法》，对人工智能治理提出了新思路。欧盟的《人工智能法》是典型的“权力—义务—责任”模式，其中权力规范对应人工智能监管机构的监管事项，义务规范与责任规范分别对应人工智能提供者、部署者（使用者）的职责与法律责任。对比借鉴之下，我国未来关于 AI 犯罪领域的立法首先宜引入权利规范。一方面权利规范的存在让义务、责任规范有了前提与基础，使用者权益保护正好是开发者、提供者负担义务的目的所在，而违反这种义务则会承担相应的法律责任。另一方面权利规范的引入还可以对权力规范形成一种制度制约，让监管机构的权力行使有了合法依据也有了行为边界。且欧盟的《人工智能法》就明确采取风险管理模式以规范人工智能活动，即根据人工智能系统可能产生的风险的强度和范围来确定规则的类型和内容，如该法案将人工智能系统可能带来的风险分为四个级别：不可接受风险、高风险、有限风险和最小风险。这种分级风险防控措施能够有效提升其准确性、透明度和稳定性，在加快人工智能发展的同时，有效地防范和控制系统风险。我国未来的立法其次应是基于人工智能活动的双重属性，明确人工智能法的科技法与应用法复式定位，在规范人工智能活动的同时最大限度地促进人工智能技术与产业的发展，并应构建重点突出的风险防范制度，既要基于“风险”本身的性质出发，考虑风险的发生概率、危害范围和程度等，也要考虑风险的事前和事后规制效益，合理配置义务和责任规则。在基于人工智能研发和应用特点的同时明确当前需重点防范的风险，动态开放地研判、防范新的潜在风险，对可能高发的涉 AI 犯罪风险的方面进行立法规范，展现法律治理设计的倾向性。最后我国应在已有制度基础和理论储备上显现出综合性立法之可行性。目前，对于人工智能数据训练阶段的数据获取问题，若刚性适用现有的个人信息保护、知识产权保护等方面的规则，可能会对人工智能的研发应用质量的提升构成法律障碍，而这类障碍在利益衡量的基础上，绝非最佳的私人权益保护方式。故仍应通过建构例外规则或特别规定等方式，协调平衡不同利益主体之间的关系，从而在不影响 AI 技术发展的同时，更好的规制 AI 深度伪造技术类犯罪。综上，我国现阶段拟定的 AI 犯罪治理法案在建构治理规则、制度等时应融通本国自主知识体系和部分国外经验资源，充分发挥显现我国当前法治发展能力的立法特色，在现有刑法规定之下，通过逐步完善司法解释的方式对法律适用过程中出现的新情况和新问题进行规制，并加强案例指导和修正案的工作，以官方形式发布专业的案例解读和相关规则释义，为司法实践提供有力的理论指导和支持，明确该技术的法律适用标准和裁判规则，用以弥补法律在 AI 犯罪治理领域的空白和漏洞。

（二）构建系统的 AI 技术监管和应对机制

AI 深度伪造涉网犯罪的治理强调建立能适应人工智能技术自主迭代和多场景切换的灵活监管机制，在宏观层面设定普适性约束底线，在微观层面将算法、数据、平台等治理对象匹配到具体场景中，设定个性化场景规则，构建普适性与个性化相结合的应对机制，强调能够针对不同时期的需求实行不同治理模式，形成“事前正向引导、事中及时调适、事后全程追踪”的动态监管治理模式，同时尊重新兴技术的发展尺度，践行包容审慎的治理原则，挑选 AI 深度伪造涉网犯罪高发领域开展治理试点测试，不断探索监管沙盒式试验性监管机制（指先划定一个范围，对在“盒子”里面的 AI 伪造技术类犯罪高发领域采取包容审慎的监管措施，同时杜绝将问题扩散到“盒子”外面，属于在可控的范围之内实行容错纠错机制，并由监管部门对运行进行全过程监管，以保证测试的安全性并作出最终的评价），讲究治理手段软法先行，充分引导与保障 AI 新兴技术发展的同时，建立系统而灵活的监管机制以应对例如“AI 去衣”等 AI 深度伪造涉网犯罪。要求从宏观层面完善技术监管机制，重点关注“AI 去衣”伪造技术的滥用和扩散问题，针对性的构建系统的监管制度规则，故相关监管部门应针对“AI 去衣”伪造技术的网络传播特点，加强网络安全防护的技术工作。例如通过加强网络安全监测、预警和应急处置技术处理能力，及时发现和处置非法 AI 图片或加大对非法 AI 图像构成的关键信息和要素进行筛选和预警，以限制此类非法 AI 图像的传播等。而监管技术部门应完善数据运算流程，借助大数据技术进行分类关联，根据此类犯罪的传播路径、关系以及相关软件介质进行特点分析，进而建立数据模型，通过系统的算法和模型精准追查制作“AI 去衣“图的非法软件和嫌疑人，并可在此基础上总结分析犯罪趋势，利用大数据的关联性分析、聚类分析等技术，针对此类犯罪的数据特征建立起高效预警机制。最后应由通信管理部门牵头、公安机关配合，严格落实网络实名和备案溯源等制度，重点针对社交媒体网络账号、图片制作软件账号采取有效监管措施，坚决杜绝“实名不实人”的问题。并通过综合采用多种生物特征识别、认证用户身份等方式，大力整治恶意批量注册账号、冒用、盗用他人账号信息等非法行为，确保注册信息真实准确、可追溯。

根据洛卡德交换原理可知，犯罪的过程实际上是一个物质交换的过程，犯罪人作为一个物质实体在实施犯罪的过程中总是会与其他的物质实体发生接触并产生互换关系。因此，犯罪案件中物质交换是广泛存在的，是犯罪行为的共生体，即使此类“AI 去衣”犯罪具有隐蔽性特征，但只要作案人实施了“AI 去衣”犯罪行为，便会留下可视化信息记录。例如图片等信息的传输必定依赖传输协议，而协议与传播者 IP 地址又存在一一对应的关系，若从 IP 地址入手，分析流量端口的 IP 地址，便能确定所在 IP 的位置区段，进而侦查人员便可以取得开通网络的用户信息，并分析其与犯罪嫌疑人的关系等。故监管部门应将大数据侦查中的异常数据信息与互联网基本规则融合，根据嫌疑人传播非法 AI 图片所必须经过或使用的工具、路径等进行严格排查和追踪，使侦查活动从虚拟空间延伸至现实空间，从而锁定犯罪嫌疑人。

（三）促进“AI 去衣”犯罪防控技术的研发与创新

由于大量“AI 去衣”软件盛行，此类

犯罪已可实现自动化，犯罪人通过 AI 软件大量生产“AI 去衣”图并进行广泛传播时，传统的防治技术已无法做到有效限制。故首先便应推动技术的创新，开发反“AI 去衣”技术进行技术反制。研发能够检测和识别“AI 去衣”伪造图像的技术，通过图像分析、机器学习等手段，对图像中的异常特征进行识别和标注，从而有效遏制该技术的滥用。同时提升图像识别技术的准确性和效率，通过引入更先进的算法和模型，提高对伪造图像的识别能力，确保能够及时发现并处理相关违法行为。其次，要针对“AI 去衣”伪造技术的网络传播特点，加强网络安全防护技术，例如通过加强防火墙、入侵检测系统、数据加密等技术手段，构建坚固的网络安全防线，防止伪造 AI 图像在网络上的大幅传播和扩散等。再者，要针对此类 AI 伪造技术产品展开对特定网络流量的监控和分析，及时发现并阻断异常的图像制作和上传活动。最后，相关监管部门应联合国家重点实验室等单位，开展“AI 去衣”图像识别与裸露图片检测技术安全测评，升级安全保护措施和相关图像算法，并覆盖即时通信、网络直播、社交媒体等重点 App，以快速发现并处置有关的非法 AI 图像信息。在上述措施的基础上，还应在宏观层面鼓励和支持科研机构、高校和企业等各方加强合作与交流，通过加大研发投入、引进优秀技术人才等方式，共同推动促进针对“AI 去衣”犯罪的防控技术研发与创新。

（四）加强 AI 伪造技术的社会共治与教育

随着 AI 技术的不断发展，已经不只是前沿的科技从业者需要了解相关 AI 知识，而是生活在当今世界的每一个人都需要熟悉相关 AI 的运作规律和使用方法。目前“AI 去衣”等深度伪造技术的犯罪影响已上升到社会层面，随时都有可能发生在我们身边，了解和使用 AI 技术已成为当下每一个公民的必修课。因此，为应对“AI 去衣”等此类人工智能伪造技术，加强社会共治和公众教育是必不可少的一环。宏观表现为：加强政府、企业、社会组织等多方面的合作和协调，形成共同打击此类犯罪的合力。具体可以从以下几个方面入手：第一，开展专项宣传活动。如政府可举办类似讲座、座谈会等形式的官方活动，向公众普及“AI 去衣”伪造技术的社会危害并传授相关防范知识。第二，加强青少年网络素养教育。如司法工作人员或是教育工作者可在学校、社区等场所开展网络素养教育活动，加强青少年的网络安全教育和思想道德水平，引导青少年树立正确的网络道德观念和行为习惯。第三，鼓励公众举报违法行为。据统计，很大一部分“AI 去衣”犯罪的受害者是因为受到威胁，害怕自己的“AI 去衣”图被进一步传播而选择隐忍，从而使得犯罪人逍遥法外。政府应通过设立举报奖励制度、保护举报人隐私等措施建立起有效的举报机制，让公众相信法律的力量，以鼓励公众积极举报此类犯罪。同时，要加强对举报线索的核实和处理工作，确保能够及时查处相关违法行为，及时消除犯罪对受害者的消极影响。第四，构建多方参与的协同治理体系。这就要求社会各领域分工明确，协同治理此类 AI 犯罪，即政府应全面发挥主导作用，加强法律法规建设和监管力度；企业应主动履行社会责任，加强 AI 技术的安全管理；社会组织应积极参与社会治理，提供相关的公益服务和支持；科研机构应加强技术研发和创新，为此类犯罪的防治工作提供技术支持和保障等。

（五）促进国际间的 AI 治理合作与交流

“AI 去衣”等伪造技术犯罪往往具有跨国性的特点，较多非法图像是通过境外网站或是境外 apk 包装软件传播制作的，因此加强国际合作与交流至关重要。应通过与国际社会加强沟通与合作，借鉴和学习国外先进的防治经验和做法，不断提升我国在网络犯罪防治领域的水平和能力。就目前而言，已有较多国家正在推行针对本土 AI 发展的治理方案，比如美国用行政令促进 AI 创新和规范保护以及通过行业自律推动监管；欧盟采用分层治理、监管沙箱等措施保护 AI 产业发展；而我国则是针对生成式 AI 开展精细化管理；在标准方面，ISO(国际标准化组织) 发布了 AI 管理框架，ITU-T(国际电信联盟电信标准分局) 启动了内容真实性标准；在技术方面，多国推动数字水印、生成内容真实性检测、深度伪造检测等。综上，可见各国对 AI 技术的具体管理规定和政策都有差别，故相当一部分犯罪者就利用了国际间管理和法律的漏洞跨国大量传输“AI 去衣”图片，以逃脱制裁，因此当前亟需建立一个国际沟通、合作和协调机制来规定统一的标准以制裁此类 AI 犯罪，该机制应明确相关标准、评估体系以及运行方式等。且这个机制的制定和运行需要 AI 领域的专家、政策和法律的制定者以及各国公民的共同参与，既要考虑普遍适用的原则，又要兼顾各国的特殊情况，在寻求共识与平衡中最大限度地利用合作机制来限制此类 AI 犯罪的发展。

本研究综合分析了“AI 去衣”此类深度伪造技术的涉网犯罪问题，并提出了一系列针对性的防治策略和措施，旨在为 AI 深度伪造技术的涉网犯罪防治提供有益的探索与参考。未来，随着 AI 技术的不断发展，新的犯罪模式和手段也会不断更新，AI 深度伪造类犯罪的防治工作也必将面临更多挑战，因此我们需要持续关注 AI 深度伪造技术的最新发展动态和趋向，不断更新对应的防治措施和策略。最后，希望未来能在各方的努力下，共同构建出一个更加安全、健康、和谐的网络空间。

参考文献：

[1]王翁杰、吴洁虹．利用 AI变脸技术违法犯罪的应对对策 [J]．广东公安科技．2020．1

[2]马长山．数字法治的体系性建构——基于 2021年以来我国数字法治建设的观察分析 [J]．浙江警察学院学报．2023．1

[3]郑志峰．人工智能使用者的立法定位及其三维规制 [J/OL]．行政法学研究．2025．1

[4]李芳、刘鑫怡．欧盟人工智能立法最新动向 [J]．科技中国．2021．6

[5]张新宝、魏艳伟．我国人工智能立法基本问题研究 [J]．法制与社会发展．2024．6

[6]刘钊、林晞楠、李昂霖．人工智能在犯罪预防中的应用及前景分析 [J]．中国人民公安大学学报 (社会科学版)．2018．4

[7]吴同．针对海量数据的数字取证模型 [J]．贵州警官职业学院学报．2011．4

[8]沈浔杰．大数据时代下涉网犯罪对策研究 [J]．法制与社会．2019．13

责任编辑　张树彦